

# Layered Index-less Indexed Flash Codes for Improving Average Performance

Riki Suzuki, Tadashi Wadayama

<sup>†</sup> Department of Computer Science, Nagoya Institute of Technology  
Email: wadayama@nitech.ac.jp

**Abstract**—In the present paper, a modification of the Index-less Indexed Flash Codes (ILIFC) for flash memory storage system is presented. Although the ILIFC proposed by Mahdavi et al. has excellent worst case performance, the ILIFC can be further improved in terms of the average case performance. The proposed scheme, referred to as the *layered ILIFC*, is based on the ILIFC. However, the primary focus of the present study is the average case performance. The main feature of the proposed scheme is the use of the layer-based index coding to represent indices of information bits. The layer index coding promotes the uniform use of cell levels, which leads to better average case performance. Based on experiments, the proposed scheme achieves a larger average number of rewritings than the original ILIFC without loss of worst case performance.

## I. INTRODUCTION

The emergence of studies by Jiang, Bohossian, and Bruck [2][3] on flash codes gave rise to a new field of coding. They modeled a flash memory as a Write Asymmetric Memory (WAM) and proposed flash codes for improving the worst number of rewritings. These studies attracted the interest of a number of coding theorists and inspired subsequent research on flash codes. In 2008, Yaakobi, Vardy, Siegel, and Wolf presented flash codes based on the enhanced multidimensional construction [4], and they proposed a novel criterion, called *write deficiency*, for flash codes.

Most flash codes proposed thus far are designed to optimize the worst case performance, such as the write deficiency [2][3][4][7]. On the other hand, recently, a flash code to improve the average number of rewritable bits was reported [5] [6] [8]. The *average number of rewritable bits* is the average number of allowable bit flips between consecutive erase operations. The flash codes based on the Gray codes proposed by Finucane, Liu, and Mitzenmacher [5] exhibit excellent average performance. They also presented a method for analyzing the average number of rewritable bits, which is based on a Markov chain model constructed from the state diagram of the code and a probabilistic model for the rewriting process [5]. In the lifetime of a flash memory, it is expected that poor average performance will result in early collapse of the cells. In this respect, flash codes should be designed to improve not only the worst case performance but also the average case performance.

The Index-less Indexed Flash Codes (ILIFC) proposed by Mahdavi et al. [7] achieve excellent worst case performance. The ILIFC has been proven to provide almost optimal write deficiency. The prominent

feature of the ILIFC is that a sub-block of cells represents both the value of an information bit and the index of the bit. This feature leads to simple encoding and decoding procedures.

In the present paper, we present an improvement of the ILIFC in terms of average case performance. Since the ILIFC is designed based on the worst case performance, there is room for improving the average case performance without incorporating drastic changes in the original algorithm.

The main concept of the scheme proposed herein is the use of *layer-based index coding*. In the original ILIFC, if a sub-block represents an index, the index cannot be changed to another index until the next erase operation. If bit flips in the information vector occur according to a nonuniform distribution, the difference between cell levels tends to be large. It is not trivial to balance the cell levels if most of the sub-blocks have their own indices. The layer-based index coding enables the encoder to change the index of a sub-block, and this feature provides flexibility for adjusting the differences between cell levels.

In the present paper, the *layered ILIFC*, which uses the layer-based index coding, will be proposed, and its average case performance will be evaluated through computer simulations and a Markov chain method designed for analyzing the average case performance of variants of the ILIFC.

## II. PRELIMINARIES

In this section, some basic assumptions on flash memories and a rewriting model are introduced according to [2].

### A. Flash memory

A flash memory contains several cells that can store electrons. The charge in a cell can be increased through the injection of electrons. The process of increasing the cell level is referred to as *cell programming*. A cell can represent  $q$ -ary values. For example, flash memories with  $q = 2$  and 4 cell levels have been realized in commercial products. The erase operation is the operation to remove the charge from cells in an *erase block*, which is a set of cells. Note that the erase operation can only be applied to all of the cells in an erase block. In other words, a cell cannot be erased individually.

There is a limitation in the number of rewritings for a cell. If the number of rewritings exceeds this limit, the cell tends to operate incorrectly. Therefore, an appropriate coding for reducing the number of rewritings is required in order to lengthen the lifetime of a flash memory [2][7].

## B. Rewriting model

In this subsection, a simple rewriting model used throughout the present paper is introduced [2][7]. The *information vector*  $(v_1, v_2, \dots, v_k)$  is a binary  $k$ -tuple, which is to be stored in an erase block. The initial state of the information vector is assumed to be  $(v_1, v_2, \dots, v_k) = (0, 0, \dots, 0)$ . The binary information in the information vector is written into cells through several rewriting processes. It is assumed that a rewriting process occurs when any one bit in the information vector flips.

The *cell state vector* for an erase block is denoted by  $(c_1, c_2, \dots, c_n)$ , where  $c_j \in \{0, \dots, q-1\}$  ( $j \in \{1, \dots, n\}$ ). The symbol  $c_j$  represents the state of the  $j$ th cell. The contents of the information vector are stored in the cell state vector. The initial state of the cell state vector is also assumed to be  $(c_1, c_2, \dots, c_n) = (0, 0, \dots, 0)$ .

Only two operations to change the state of a cell state vector are allowed, namely, cell programming and the erase operation. Cell programming increases a cell level by one until the level reaches  $q-1$ . Suppose that there are two state vectors  $\mathbf{c} = (c_1, c_2, \dots, c_n)$  and  $\mathbf{c}' = (c'_1, c'_2, \dots, c'_n)$ . If  $c_j \geq c'_j$  holds for any  $j \in \{1, 2, \dots, n\}$ , then  $\mathbf{c}$  is said to be *higher* than  $\mathbf{c}'$ . These assumptions mean that the allowable state transitions of an erase block are a transition from a lower state to a higher state. An erase operation forces the cell state vector to be a zero vector.

A goal of the flash codes is to reduce the number of erase operations to be as small as possible, under the assumption that the number of bit flips of information bits is fixed.

## III. INDEX-LESS INDEXED FLASH CODES

In this section, a brief introduction of the ILIFC as reported by [7] is presented. Although an encoding and decoding process of the ILIFC consists of multiple stages, for simplicity, we herein focus only on the first stage.

Assume that an erase block contains  $n$ -cells and that these cells are divided into sub-blocks of the same size and any remaining cells. The size of a sub-block is assumed to be  $k$  satisfying  $n > k^2$ , and  $k \bmod 2 = 0$ . The number of sub-blocks is given by  $m = \lfloor n/k \rfloor$ , and the number of remaining cells is  $n - km$ . The length of the information vector is  $k$ . The key feature of the ILIFC is the representation of individual bits in the information vector. Each sub-block in an erase block represents both an index and a value of an information bit.

Let us denote the state of an erase block as  $(\mathbf{x}_1 | \mathbf{x}_2 | \dots | \mathbf{x}_m)$  where  $\mathbf{x}_i \in \{0, \dots, q-1\}^k$  ( $i \in \{1, \dots, m\}$ ) is the  $i$ th sub-block. For a sub-block  $\mathbf{x} = (c_1, c_2, \dots, c_k)$ , the weight and the parity of  $\mathbf{x}$  are defined as  $\text{wt}(\mathbf{x}) = c_1 + c_2 + \dots + c_k$ , and  $\text{parity}(\mathbf{x}) = \text{wt}(\mathbf{x}) \bmod 2$ , respectively.

The following terminology is used throughout the present paper. A state of a sub-block is said to be

- 1) *full* if all cells in the sub-block have level  $q-1$ ;
- 2) *empty* if all cells in the sub-block have level 0;
- 3) *active* if the sub-block is neither full nor empty;
- 4) *live* if the sub-block is not live.

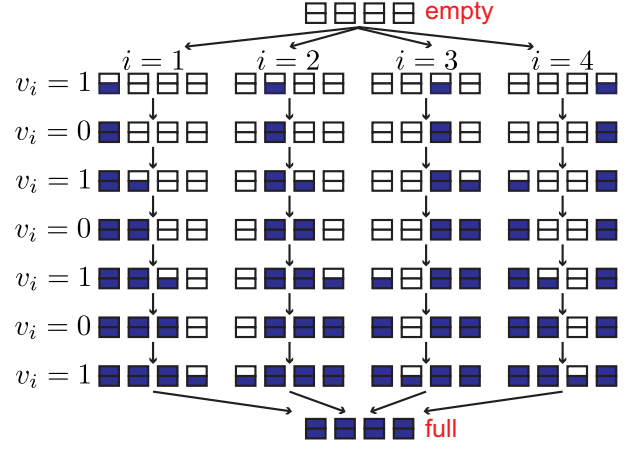


Fig. 1. Index coding for the ILIFC ( $k=4, q=3$ ). The variable  $i$  expresses the value of the index. Whenever information bit  $v_i$  turns, a state transition (an arrow) from a cell state vector to a higher cell state vector occurs.

## A. Index coding for the ILIFC

A sub-block represents a single bit in the information vector. The details are as follows. Let  $I : \{0, \dots, q-1\}^k \rightarrow \{0, 1, 2, \dots, k\}$  be an *index map* that converts a state of a sub-block into an index of the information vector. In [7], the index map is given by

$$I(\mathbf{x}) = \begin{cases} \arg \max_{i \in \{1, \dots, k\}} [c_i - c_{(i-2 \bmod k)+1}], & \mathbf{x} \neq \mathbf{0} \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

where  $\mathbf{x} = (c_1, c_2, \dots, c_k)$ . Note that  $I(\mathbf{x}) = 0$  implies that the sub-block  $\mathbf{x}$  has no index. The *bit value map* for a sub-block:  $V : \{0, \dots, q-1\}^k \rightarrow \{0, 1\}$  is given simply as  $V(\mathbf{x}) = \text{parity}(\mathbf{x})$ .

Based on the definition presented above, a sub-block  $\mathbf{x}_i$  can represent both an index  $I(\mathbf{x}_i)$  and the bit value  $V(\mathbf{x}_i)$ . At any moment during an encoding process (to be described later), the ILIFC encoder maintains the consistency as follows:  $v_{I(\mathbf{x}_i)} = V(\mathbf{x}_i)$  if  $I(\mathbf{x}_i) \neq 0$  for  $i = 1, 2, \dots, k$ , where  $(v_1, \dots, v_k)$  is the information vector. A coding that achieves this consistency is referred to as an *index coding*.

In the ILIFC, the index coding for a sub-block is carefully designed in order to avoid an early erase operation. Figure 1 illustrates the process for the case in which  $k=4$  and  $q=2$ . In Fig. 1, the variable  $i$  corresponds to the index. The position at which the level of the cells begins to rise corresponds to the index  $i$  and is consistent with the definition of  $I(\cdot)$ . Furthermore, a bit flip on  $v_i$  induces an increment of a cell value that corresponds to the arrow shown in Fig. 1. Note also that the states of two active sub-blocks cannot have the same state if the indices of two sub-blocks are different.

## B. Pseudo-code for the ILIFC encoder and decoder

The decoding and encoding algorithms for the ILIFC are shown in the form of a pseudo-code according to [7].

### Decoding map

$$(v_1, v_2, \dots, v_k) = (0, 0, \dots, 0);$$

```

for ( $j = 1; j \leq m; j = j + 1$ )
if (active( $x_j$ ))
{  $i = \text{read\_index}(x_j); v_i = \text{parity}(x_j);$  }

```

#### Encoding map

```

 $y = (y_1 | y_2 | \dots | y_m) = (x_1 | x_2 | \dots | x_m);$ 
for ( $j = 1; j \leq m; j = j + 1$ ) {
if (active( $x_j$ )  $\wedge$  (read_index( $x_j$ ) ==  $i$ ))
{ write( $y_j$ ); return( $y$ ); }
}

```

```

for ( $j = 1; j \leq m; j = j + 1$ )
if (empty( $x_j$ ))
{ write_new( $i, y_j$ ); return( $y$ ); }
return E;

```

The symbol **E** represents the erase operation. The details of the functions, such as **active**( $\cdot$ ), can be found in [7].

#### IV. LAYERED ILIFC FOR IMPROVING AVERAGE CASE PERFORMANCE

In this section, we propose a modified ILIFC to improve the average case performance. A significant difference between the original ILIFC and the modified ILIFC is the index coding for a sub-block. The modified ILIFC introduces the concept of the *layer* in its index coding. The layer in a sub-block enables us to reset a sub-block, which means that an index embedded in a sub-block can be changed to another index. This flexibility promotes the uniform use of cell levels in an erase block and leads to an improvement in the average case performance.

##### A. Outline of Layered ILIFC

In the following, for simplicity, we assume that an erase block contains exactly  $n = k^2$  cells.

In the encoding process of the original ILIFC, the index of a sub-block is fixed until the next erase operation occurs. Thus, if several bits in the information vector are frequently rewritten, the sub-blocks corresponding to these bits tend to become full in early phase. In other words, local rewritings in the information vector induce an imbalance of cell levels between the sub-blocks. This imbalance tends to cause an early erase operation. In order to overcome this problem, maintaining the balance of cell levels for any rewriting sequence is crucially important.

In the modified ILIFC, the cells in a sub-block are programmed from the bottom layer to the top layer. If a layer is filled (or programmed), we may alter the index of the sub-block. In the following, we refer to the proposed modified scheme as the *layered-ILIFC* (L-ILIFC).

##### B. Index coding for Layered ILIFC

In the following, we assume that  $k$  is an even number. Suppose that a cell  $c_i$  has the value  $l \in \{0, \dots, q-1\}$ . In such a case, the cell  $c_i$  is said to be in layer  $l$ . In other words, cell levels are divided into  $q$ -layers.

Next, we introduce some additional considerations regarding the state of a sub-block. If all of the cells in a sub-block

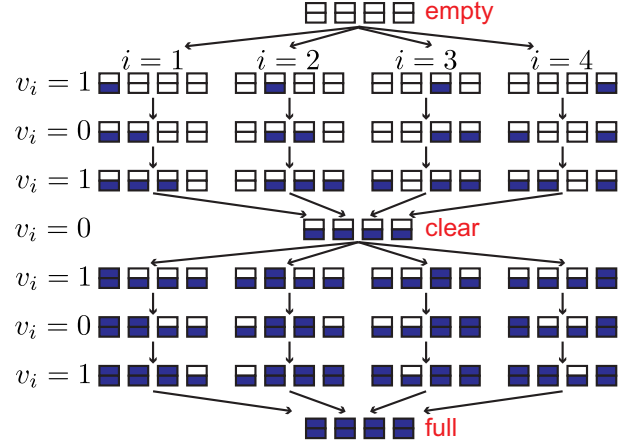


Fig. 2. Index coding for the layered ILIFC ( $k = 4, q = 3$ ); The variable  $i$  expresses the value of the index. Whenever information bit  $v_i$  turns, a state transition (an arrow) from a cell state vector to a higher cell state vector occurs.

belong to the same layer  $l \in \{0, 1, \dots, q-2\}$ , then the sub-block is said to be *clear*.

We first explain the concept of index coding using the following example. The details of the index coding will be presented in Subsection IV-C. Figure 2 presents the details of the index coding for  $k = 4$  and  $q = 2$ . Initially, the cell is empty (empty boxes at the top in Fig. 2). A single bit change in the information vector corresponds to the arrows emerging from the empty state. An index  $i$  (from 1 to 4) is written in a sub-block by incrementing the value of the  $i$ th cell, and the parity of the sub-block becomes one.

It is readily observed that two more changes in the information vector transforms the state of the sub-block into the clear state. Namely, all of the cells belong to the same layer. A clear sub-block has no index. Note that we can write any index into a clear sub-block, as shown in Fig. 2.

The index map for this index coding is, thus, given by

$$I'(\mathbf{x}) = \begin{cases} 0, \forall i, j \in \{1, \dots, k\}, c_i = c_k, \text{ (i.e., } \mathbf{x} \text{ is clear)} \\ \arg \max_{i \in \{1, \dots, k\}} [c_i - c_{(i-2 \bmod k)+1}], \text{ otherwise.} \end{cases} \quad (2)$$

Due to the assumption that  $k$  is even, any clear sub-block has a value of zero. This is because the sum  $c_1 + \dots + c_k$  becomes an integer multiple of  $k$ . Therefore, there is no problem for a sub-block to forget both its index and the value when it becomes a clear sub-block.

##### C. Details of the layered ILIFC

The pseudo-code of the encoding map of the layered ILIFC is shown below.

#### Encoding map

```

 $y = (y_1 | y_2 | \dots | y_m) = (x_1 | x_2 | \dots | x_m);$ 
for ( $j = 1; j \leq m; j = j + 1$ ) {
if ( $(\neg \text{clear}(x_j)) \wedge (\text{read\_index2}(x_j) == i)$ )
{ write2( $y_j$ ); return( $y$ ); }
}

```

```

for ( $l = 0$ ;  $l \leq q - 1$ ;  $l = l + 1$ ) {
for ( $j = 1$ ;  $j \leq m$ ;  $j = j + 1$ ) {
if ( $\text{clear}(x_j) \wedge (\text{read\_layer}(x_j) == l)$ )
{ $\text{write\_new2}(i, y_j)$ ;  $\text{return}(y)$ ;}
}
}

```

**return E;**

The clear sub-blocks in an erase block can be considered to be a resource for achieving flexibility. Assume that the encoder must write the index value  $i$  in a sub-block. If there is no active sub-block with the index value  $i$ , then the encoder searches for a clear sub-block in the lowest layer. Note that the encoder of the original ILIFC can choose a sub-block to write the index value  $i$  only from among the empty sub-blocks. In other words, the existence of clear sub-blocks and this search method reduces the nonuniform use of the cell levels.

The details of the functions used in the pseudo-code are given as follows. The function **clear**( $x$ ) returns true if  $x$  is clear; otherwise the function returns false. The function **read\_layer**( $x$ ) returns the layer index of  $x$  if  $x$  is clear; otherwise the function returns  $-1$ .

The function **read\_index2**( $x$ ) is defined as

$$\text{read\_index2}(x) = I'(x).$$

The function **write\_new2**( $i, x$ ) is a function to write the index value  $i$  into the clear sub-block  $x$ . This function simply increments the value of  $c_i$  by one. The function **write2**( $x$ ) changes the value of the active sub-block  $x$  by incrementing  $c_{(i+\text{wt}(x)) \bmod k}$  where  $x = (c_1, \dots, c_k)$ . This rule of the change in the value of a sub-block corresponds to the arrows in Fig. 2.

#### D. Worst case performance

The worst number of possible rewritings between a consecutive erase operation for the proposed scheme is same as that for the original ILIFC. This is because both schemes share the same worst case events. A worst case event occurs when the cell vector contains one full sub-block and  $k - 1$  active sub-blocks of cell level 1.

### V. MARKOV CHAIN METHOD

In this section, we briefly review the Markov chain method for average case analysis [5] and then apply the Markov chain method to the analysis on the ILIFC and the Layered ILIFC.

The concept of the Markov chain method as follows. We first construct a state transition diagram for a flash code to be analyzed. A state of the diagram is an allowable state of the cell state vector. Each edge (i.e., state transition) has its own probability, which is determined based on the probabilistic model of rewriting sequences for the information vector. The state transition diagram naturally defines a Markov chain corresponding the pair of the flash code and the probabilistic model.

The steady state probability of the Markov chain can be obtained by solving a simultaneous linear equation system defined by the transition probability matrix. The average

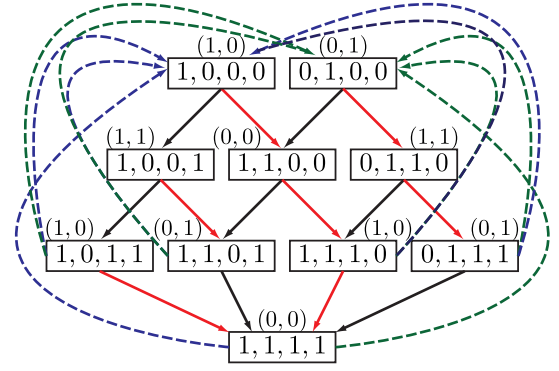


Fig. 3. An example of the Markov chain for ILIFC ( $n = 4, k = 2, m = 2, q = 2$ )

TABLE I  
AVERAGE PROBABILITY OF OCCURRENCE OF ERASE OPERATIONS OF THE LAYERED ILIFC ( $n = 4, k = 2, q = 4$ ).

	$P_1 = 1/2$	$P_1 = 1/5$
Computer simulation	0.091006	0.095430
Markov chain	0.091006	0.095431

The probability  $P_1$  is the probability such that the first bit of the information bit is flipped.

number of possible rewritings is directly derived from the steady state probability.

Figure 3 illustrates the state transition diagram of the ILIFC with the parameters  $n = 4, k = 2$ , and  $q = 2$ . The binary 4-tuples in the boxes are the states of the cell state vector ( $n = 4$ ), and the binary 2-tuples in parentheses represent the states of the information vector. The arcs connecting the boxes denote possible state transitions. The dotted arcs correspond to the state transitions that induce an erase operation.

The average probability of the occurrence of an erase operation is given by the sum of the probabilities of the state transitions corresponding to the dotted arcs. This probability can be calculated from the steady state probability. The average number of possible rewritings is given as the inverse of this probability [5].

If the state space of the ILIFC or the Layered ILIFC is not so large, then this Markov chain method is useful to obtain an accurate estimate. Table I compares the average probability of an occurrence of an erase operation obtained through computer simulation and the Markov chain method. We have confirmed that the values obtained by the Markov method and the computer simulations are in reasonable agreement.

### VI. RESULTS OF COMPUTER EXPERIMENTS

We performed computer experiments to compare the Layered ILIFC and the original ILIFC in terms of the number of possible rewritings between two consecutive erase operations.

The assumptions made in the simulations are as follows. In unit time, only one bit in the information vector flips, which induces a state change in the cell state vector. The flipped bit is chosen according to the uniform distribution from 1 to  $k$ . The same pseudo-random sequences are used for both schemes

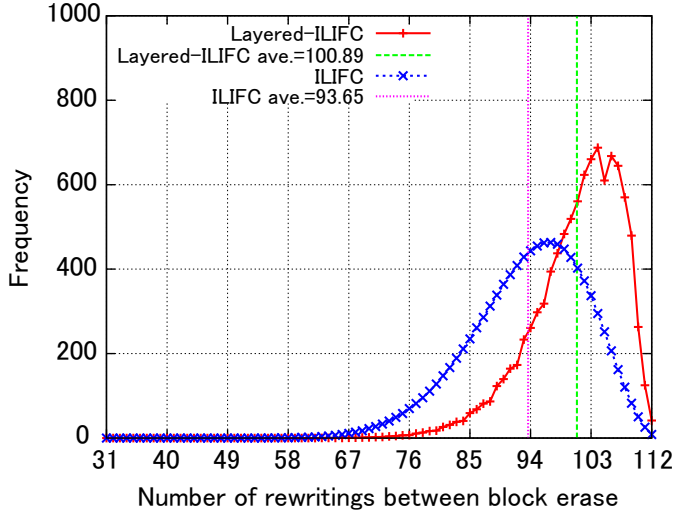


Fig. 4. Frequency of the number of rewritings between the block elimination of ILIFC and layered ILIFC ( $n = 16$ ,  $k = 4$ ,  $m = 4$ ,  $q = 8$ ).

in order to realize a fair comparison. In the simulation, the information vector and the cell state vector are initialized after a block erase operation.

Figure 4 presents histograms for the number of rewritings between consecutive erase operations. The horizontal axis represents the number of rewritings, and the vertical axis denotes the frequency of the number of rewritings observed in the experiment. In an experiment, rewriting operations are iterated until the number of erase operations becomes  $10^4$ -times. The parameter settings are as follows. The number of cells in an erase block is  $n = 16$ , and each cell can take a value from 0 to 7 (i.e.,  $q = 8$ ). The length of the information vector is  $k = 4$ , and thus  $m = 4$ .

It is readily observed that the histogram curve of the layered ILIFC indicates a desirable behavior of the layered ILIFC. If the number of rewritings is not so large (i.e., smaller than 94), then the layered ILIFC provides a much smaller frequency than the ILIFC. The figure also includes the average number of rewritings. The layered ILIFC yields 100.89 rewritings, and the original ILIFC yields 93.65 rewritings. On average, approximately seven more rewritings are possible with the layered ILIFC.

The Markov chain method presented in Section V provides accurate values of the expected number of rewritings for the layered ILIFC and the ILIFC.

Figure 5 shows the trade-off curves between the code rate versus the average number of rewritings. The code rate is defined as  $R = k/n$ . The parameters  $k = 2$  and  $n = 4$  are assumed. The layered ILIFC provides a better trade-off compared with the original ILIFC. This indicates that the proposed scheme has better average case performance than that of the original scheme.

## VII. CONCLUSIVE SUMMARY

The original ILIFC is an excellent flash code having near optimal write deficiency. In the present paper, we revealed

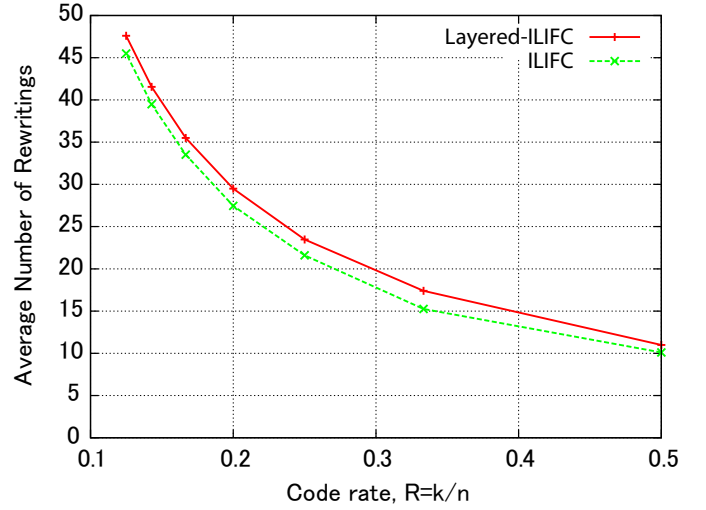


Fig. 5. Average number of rewritings between consecutive erase operations of ILIFC and layered ILIFC ( $k = 2$ ,  $q = 4$ ).

that a simple modification of the index coding promotes the uniform use of the cell level. In addition, we demonstrated that the average performance of the ILIFC can be further improved without loss of worst case performance. The Markov chain method for the ILIFC may be a useful tool for optimizing the detail of the algorithm in terms of average performance.

## REFERENCES

- [1] R. L. Rivest and A. Shamir, "How to reuse a 'write-once' memory," *Information and Control*, vol. 55, pp. 1-19, 1982.
- [2] A. Jiang, V. Bohossian, and J. Bruck, "Floating codes for joint information storage in write asymmetric memories," *Proc. IEEE International Symposium on Information Theory*, pp. 1166-1170, Nice, France, June 2007.
- [3] A. Jiang and V. Bruck, "Joint coding for flash memory storage," *Proc. IEEE International Symposium on Information Theory*, pp. 1741-1745, Toronto, Canada, July 2008.
- [4] E. Yaakobi, A. Vardy, P. H. Siegel, and J.K. Wolf, "Multidimensional flash codes," *Proc. 46-th Allerton Conference on Communication, Control and Computing*, pp. 392-399, Monticello, IL, Sept. 2008.
- [5] H. Finucane, Z. Liu, and M. Mitzenmacher, "Designing floating codes for expected performance," *Proc. 47th Allerton Conf. (Illinois)*, pp. 1389-1396, Sept. 2008.
- [6] H. Finucane and M. Mitzenmacher, "Worst-case and average-case floating codes for flash memory," *Harvard University*, 2009.
- [7] H. Mahdavi, P. H. Siegel, A. Vardy, J. k. Wolf, and E. Yaakobi, "A nearly optimal construction of flash codes," *CoRR*, abs/0905.1512, 2009.
- [8] H. Kamabe, "Floating codes with good average performance," *SITA2009*, pp. 856-861, 2009.